Group Decision Making with Dispersion in the Analytic Hierarchy Process

Natalie M. Scala, Jayant Rajgopal, Luis G. Vargas, Kim LaScola Needy

Journal Website: https://link.springer.com/article/10.1007/s10726-015-9445-7

Article DOI: https://doi.org/10.1007/s10726-015-9445-7

Group Decision Making with Dispersion in the Analytic Hierarchy Process

ABSTRACT

With group judgments in the context of the Analytic Hierarchy Process (AHP) one would hope for broad consensus among the decision makers. However, in practice this will not always be the case, and significant dispersion may exist among the judgments. Too much dispersion violates the principle of Pareto Optimality at the comparison level and/or matrix level, and if this happens, then the group may be homogenous in some comparisons and heterogeneous in others. The question then arises as to what would be an appropriate aggregation scheme when a consensus cannot be reached and the decision makers are either unwilling or unable to revise their judgments. In particular, the traditional aggregation via the geometric mean has been shown to be inappropriate in such situations. In this paper, we propose a new method for aggregating judgments when the raw geometric mean cannot be used. Our work is motivated by a supply chain problem of managing spare parts in the nuclear power generation sector and can be applied whenever the AHP is used with judgments from multiple decision makers. The method makes use of principal components analysis (PCA) to combine the judgments into one aggregated value for each pairwise comparison. We show that this approach is equivalent to using a weighted geometric mean with the weights obtained from the PCA.

KEYWORDS: Group decisions; Analytic Hierarchy Process; geometric mean; geometric dispersion; principal component analysis; weighted geometric mean

1. INTRODUCTION

This paper addresses aggregation of multiple judgments in the context of the Analytic Hierarchy Process (AHP). Specifically, it looks at the case where there is disagreement within the group of decision makers, and the members of the group are either unwilling or unable to revise their judgments. The geometric mean (GM) of the pairwise comparisons of the group is the traditional means for aggregating group judgments; significant disagreement is manifested in the form of excessive dispersion around the GM. The explicit consideration of dispersion in the context of the AHP is relatively new. Saaty and Vargas (2007) introduced a dispersion test for group judgment aggregation to ensure that variability around the geometric mean is small enough so that all comparisons remain homogeneous. In this paper we address situations where the test results indicate that this variability is not sufficiently small. We develop a technique for aggregating pairwise comparisons through the use of the weighted geometric mean, with principal components analysis (PCA) being used to calculate weights for every decision maker. Our methodology is motivated by a supply chain problem in the nuclear power generation sector, where group pairwise comparisons were elicited in the context of the AHP (Scala, Rajgopal, and Needy 2014). The PCA-based approach to aggregation that we develop is illustrated using data from this case study. However, aggregation of group judgments is common in the AHP, and our work is relevant to any situation where the conditions outlined above hold, or more generally, when we wish to give different weights to the different judges.

We begin with a brief discussion of the AHP, especially in the context of group decision making. We then discuss the dispersion test of Saaty and Vargas (2007) for aggregation of group judgments and motivate the need for the PCA-based approach for aggregation that is developed in this paper. Section 3 provides the development and details of our approach; Section 4 further examines its properties and behavior; and Section 5 illustrates it with an example using the nuclear supply chain data discussed above.

2. BACKGROUND AND LITERATURE

2.1 The AHP, Group Decision Making, and the Dispersion Test

The Analytic Hierarchy Process (AHP) is a structured methodology for decision-making in a complex environment. It is a multiattribute approach based on three principles: *decomposition*, *measurement* and *synthesis*. Decomposition constructs a hierarchical network with the goal of the analysis at the top, criteria and subcriteria in the middle tiers, and a bottom tier of decision alternatives. Measurements are in the form of pairwise comparisons between each set of criteria at a level (or alternatives at the lowest level) that are made with respect to the criteria at the next higher level or overall goal. Synthesis composes these comparisons and maps them onto a unidimensional scale for identifying the best alternative.

As an example, consider a problem that has a three-level hierarchy, with the goal at the top, three criteria in the middle, and five decision alternatives at the bottom. Then a set of ten pairwise comparisons between the five alternatives would be conducted three times; once each with respect to each of the three criteria. A judgment resulting from a pairwise comparison belongs to an absolute ratio scale and indicates how much more important the first alternative is than the second, with respect to the criterion in question. These judgments satisfy the *reciprocal property*, i.e., if a is x times more preferred than b, then b should be $1/x$ times less preferred than a. The pairwise comparisons are then synthesized through the use of linear algebra, and priorities for each alternative are computed. The priorities are also normalized to sum to one. The alternative with the highest priority value is then said to be the preferred decision. For further description of the AHP, the reader is referred to Saaty (1980; 1990; 2013). A review of theoretical advancements in the AHP since its inception can be found in Ishizaka and Labib (2011a).

Often it is appropriate to have more than one decision maker perform the pairwise comparisons in the AHP analysis. This allows for multiple points of view and the knowledge of multiple subject matter experts to play a role in the final decision. These multiple decision makers can work separately, be spread out over multiple geographic locations, or be in a centralized setting. Their analysis leads to individual judgments which can then be aggregated into group judgments in order to obtain a more robust set of comparisons. Traditionally, for a given pairwise comparison, the GM of the judgments across all decision makers is computed and used as the aggregate judgment for the group. It has been shown by Aczél and Saaty (1983) to be the only mathematically valid synthesis method for AHP. In addition to homogeneity and unanimity, it also maintains the reciprocal property; all three properties are axioms of the AHP. The GM is the mathematical equivalent of consensus if all judges are considered equally important/reliable, otherwise a weighted geometric mean (WGM) is reasonable (Aczel and Alsina 1987). There are many instances in the AHP literature where the geometric mean has been used to aggregate group judgments. Examples include Armacost, Hosseini, and Pet-Edwards (1999), Liberatore and Nydick (1997), Lai, Wong, and Cheung (2002), and Wei, Chien, and Wang (2005). However, none of these address the dispersion of judgments; a geometric mean was used to aggregate judgments without explicitly testing for dispersion.

The AHP has been widely used as a multi-criteria decision-making technique. Examples include supplier selection in supply chains (Liu and Hai 2005; Ramanathan, 2007), ERP system selection (Wei, Chien, and Wang 2005) and warehouse management (Korpela, Lehmusvaara, and Nisonen 2007). In some cases, multiple decision makers are considered; in other cases, a single decision maker provides the comparisons. A recent review of applications of AHP in operations management, including decisions related to operations strategy, process and product design, planning and scheduling resources, project management, and managing the supply chain, can be found in Subramanian and Ramanathan (2012).

While the best case with group judgments is broad consensus among the decision makers, in practice this does not always occur. Saaty and Vargas (2007) address aggregation in conditions where a consensus cannot be reached and significant dispersion exists among the judgments and show that the geometric mean cannot automatically be used in all instances. Specifically, one needs to look at the amount of dispersion around the geometric mean of the judgments. Too much dispersion violates the principle of Pareto Optimality at the comparison level and/or matrix level, which is crucial to the

AHP. If this happens, the group may be homogenous in some comparisons and heterogeneous in others.

In their paper, Saaty and Vargas (2007) develop a formal statistical test for dispersion, which is designed to determine if the observed variance in the set of group judgments for a given pairwise comparison is typical, given the group's behavior. The authors show that as the number of decision makers (m) increases, the geometric dispersion of the comparisons, a measure introduced by Saaty and Vargas (2005), tends to follow a gamma distribution whose shape and scale parameter values can be determined for any value of m. Thus, their statistical test determines the probability (p) of randomly observing a geometric dispersion of the group judgments that is less than or equal to the calculated (sample) value for the group. If this probability is sufficiently small, then it would indicate that there is a high probability ($1-p$) of random geometric dispersions that are at least as large as the calculated value, thus implying that the observed dispersion is not unusually large and the geometric mean can be used. Conversely, if the p-value is large, the judgments are deemed to fail the dispersion test, and it would be inappropriate to use the geometric mean to aggregate the judgments across that particular pairwise comparison. Further discussion and details on how to perform the dispersion test are given in Saaty and Vargas (2007).

When dispersion is large and a set of pairwise comparisons cannot be aggregated directly, the literature (Basak 1988; Dyer and Forman 1992; Basak and Saaty 1993; Aczél and Saaty 1983; Aczél and Alsina 1987; Saaty and Vargas 2007) does not provide any formal alternative; rather, it directs the decision makers to "work together" to reach consensus, with judgments being revised or reconsidered. If the decision makers choose to revise their judgments, and the corresponding set of pairwise comparisons is reasonably consistent in the AHP, then those revised judgments may be substituted for the decision maker's original judgments. Another dispersion test should be performed, and if the test is passed, then the judgments can be aggregated.

In practice though, this process of returning to the decision makers is often not feasible from a logistical or geographical standpoint. While group meetings at the same place and time can perhaps be obviated due to technology advancements (as noted by Huang, Liao, and Lin, 2009), consensus is best achieved by having the decision makers collectively meet in the same location for a face to face discussion. However, this might not be feasible. Moreover, the decision makers may simply not want to revise their judgments, or the revised judgments might still not pass the dispersion test. This could be possible in survey situations for example, where the decision makers are geographically dispersed or responding without group interaction. In this case, the AHP literature has not addressed, to our knowledge, the question of how one should proceed with group judgment aggregation in light of the development of Saaty and Vargas's (2007) dispersion test.

This paper develops a new method for aggregating judgments when there is large geometric dispersion and the decision makers are either unwilling or unable to revise their judgments. Our method makes use of principal components analysis (PCA) to combine the judgments into one aggregated value for each pairwise comparison by addressing the inherent variability among the different judges. We show that this approach is equivalent to using a weighted geometric mean with the weights coming from the PCA. The weighted geometric mean preserves the unanimity, homogeneity and the reciprocal properties of the AHP (Aczél and Alsina 1987). It also serves the role of giving different weights to the different judges in the group, because they are presumably not all consistent to the same degree. This paper also addresses the general issue of objectively determining exactly *how* the decision makers' weights for the weighted geometric mean should be selected (regardless of whether the judgments do or do not pass the dispersion test).

2.2 Group Aggregation

Prior to the development of the dispersion test by Saaty and Vargas (2007), a geometric mean was used for aggregation regardless of the amount of dispersion between group judgments. However, to our knowledge, no objective alternatives have been proposed for situations when the dispersion test fails. Indeed, as Ishizaka and Labib (2011a) note, a vast majority of papers in the AHP use the method as it was first described by Saaty and tend to ignore current developments. Basak (1988), and Basak and Saaty (1993) examined a related problem in group aggregation. In their work the decision makers are divided into separate groups. The authors take a statistical approach, and present a method where the groups are tested for homogeneity to see if some of the groups can be combined. This method assumes there are a large number of dispersed decision makers and uses maximum likelihood estimators with the likelihood ratio test. It also assumes that the judgments of the decision makers are distributed lognormally, which may or may not be true for a given group of decision makers' judgments. However, the authors do not address how the decision makers are initially grouped together. More importantly, while their method identifies when groups can be pooled and when they cannot, it does not provide guidelines on what to do in the latter case.

Huang, Liao, and Lin (2009) developed a group aggregation approach through preferential differences and ranks, but do not mention or utilize the dispersion test, even though the authors discuss divergent opinions. Their method does not use the geometric mean for judgment aggregation, nor do they mathematically prove that their method does not violate the axioms of the AHP. Escobar and Moreno-Jimenez (2007) developed an aggregation method that combines the AHP and Borda count methods. They argue that their method compares with row geometric mean aggregation, but they too do not check for dispersion among their decision makers nor whether the aggregation is valid. Arias-Nicolás, Pérez, and Martín (2008) developed an aggregation method using logistic regression and group meetings. However, they do not consider dispersion in their aggregation scheme. Pedrycz and Song (2011) develop an aggregation method using granular matrices, and define consensus using the inconsistency ratio of each decision maker. Like the other papers mentioned, they also do not explicitly consider dispersion. Xu and Yager (2010) develop a power ordered weighted geometric operator for aggregation. The challenge with their results is that decision makers can be individually consistent but can be dispersed at the group level, and once again dispersion is not considered. Finally, van den Honert (2001) proposed a group decision aggregation technique using a combination of AHP and the Simple Multi-Attribute Rating Technique (SMART) but this method also does not specifically address dispersion. The author points out that a limitation of his work is that decision makers may form coalitions, which would reduce to equal weight for every decision maker in an extreme case.

There has been some prior work in aggregation methods that assign weights to judges. For example Cho and Cho (2008) assign weights based on the inconsistency ratio using Taguchi's loss function from quality control theory. The authors examine the individual consistency of each decision maker and create an exponential evaluation reliability function to derive weights. A test for the overall dispersion of judgments is not included in this method; in theory, the decision makers could be individually consistent but still dispersed when considered as a group. Furthermore, the author notes that each group would have its own evaluation reliability function, which may or may not be exponential; methods and distributions for deriving other functions are not included in the paper.

Another example is Ishizaka and Labib (2011b), who develop the Group Analytic Hierarchy Process Ordering (GAHPO) method. In this approach, the stakeholders are incorporated into a level of the hierarchy. The decision makers then develop weights for themselves by working through the pairwise comparisons of the hierarchy. Each decision maker reserves a veto power to refute his/her weight as developed by the group. Dispersion of group judgments is not formally addressed in this method. Clearly, the veto powers can introduce politics into the process, and the authors do not formally address how to deal with this issue or with a strong decision maker who does not agree with the majority. Furthermore, scenarios may exist where the decision makers do not personally know each other or may not be very familiar with each other. As a result, the quality of the pairwise comparisons of the decision maker may suffer; without knowledge of each decision maker, his/her value or relevance may not be properly reflected. The authors note that voting methods or methods where every decision maker evaluates the same hierarchy are best for synergistic groups and not for cases where the group is a collection of strong individual opinions.

A third example is a paper by van den Honert (1998) which uses the multiplicative AHP with means and variances of pairwise comparisons for group preference and does not consider dispersion. The author discusses concerns with group preference in that once judgments are aggregated, the group may not be pleased with the overall result, resulting in uncertainty. However, the concern for uncertainty may be avoided with a test for dispersion; if the judgments are dispersed, they should be revised or addressed instead of declaring uncertainty. Furthermore, Vargas (1997) presents a counterexample to the validity of the multiplicative AHP and raises issues with both its eigenvector and aggregation of priorities versus judgments in a group setting. Therefore, we argue that the multiplicative AHP is not the best approach to use.

2.3 Pareto Optimality

Fundamentally, the principle of Pareto Optimality implies that if all individuals prefer A to B then so should the group. The principle is violated when the group may be homogeneous in some paired comparisons and heterogeneous in others. Ramanathan and Ganesh (1994) explored two methods of combining judgments in hierarchies, but both methods violated Pareto Optimality for pairwise comparisons, and as a result, the authors incorrectly conclude that the geometric mean violates Pareto Optimality. Forman and Peniwati (1998) argue that this claim is irrelevant, because in group aggregation, we are no longer concerned with individual priorities. The authors also note that in group decision making, for the good of the group the experts give up their preferences, with the group becoming a new individual. Furthermore, each decision maker may not provide pairwise comparisons for each portion of the AHP hierarchy. Van den Honert and Lootsma (1996) show that group judgment aggregation does satisfy Pareto Optimality and that Ramanathan and Ganesh's proposition is not practical. Ishizaka and Labib (2011a) note that the concept of dispersion further downplays Ramanathan and Ganesh's (1994) claims, arguing that too much dispersion implies a heterogeneous group and further work with the decision makers is needed to achieve homogeneity and consensus among the group. Saaty and Vargas (2012) prove that a social welfare function exists that satisfies multiple conditions including Pareto Optimality when using the geometric mean to combine individual judgments, further challenging earlier claims. As Saaty (2013) notes, this proof justifies mathematically that individuals can combine judgments into a representative group judgment. He also argues that a mathematical way of combining judgments is preferable to arbitrarily declaring consensus.

This paper seeks to provide another step in group aggregation research, when the AHP is applied to a decision-making problem. Specifically we address the case where decision makers are dispersed and unwilling or unable to revise their judgments, and when the dispersion test outcome is not favorable. The proposed approach does not violate the axioms of the AHP, as it generalizes to aggregation through a weighted geometric mean. In the broader context of group aggregation, it also provides a method for determining weights for decision makers when using a weighted geometric mean. Next, we demonstrate this method.

3. APPROACH

We begin this section with a brief background discussion of PCA, and then discuss how it both relates to the AHP and yields a methodology for synthesizing judgments. We also provide an empirical study that shows that the PCA-based approach for aggregation generalizes to the GM approach when the decision makers are not dispersed.

3.1 Overview of PCA

Principal components analysis is a statistical technique that uses an orthogonal linear transformation to transform a set of (most likely correlated) variables into a smaller set of uncorrelated variables (Dunteman 1989). In essence, PCA attempts to reduce the dimensionality of a data set by distilling it down to the components that capture the majority of the variability associated with the original data set. The procedure addresses the variance/covariance structure of the data and uses the eigenvectors of the covariance matrix to transform the data to a new coordinate system in which the original data is rotated on its axes such that maximum variance of any projection of the data lies along the first coordinate (the first principal component), the second largest variance along the second coordinate (the second principal component), and so on. Thus, if we start with a set of n observations in m variables, PCA reduces the original data set to n observations on k *components* that capture a large proportion of the total variance in the original data set. Principal components are used for data reduction and interpretation and are typically employed in large analyses, revealing relationships in the data that might not originally be evident. The method has been widely applied in practice, including areas such as biology, medicine, chemistry, and geology (Dunteman 1989). Further details on principal components can be found in Johnson and Wichern (2007), Anderson (2003), Jolliffe (2002), and Dunteman (1989).

3.2 Principal Components and the AHP

Now consider group judgments in the context of the AHP. Each of the n pairwise comparisons made by the set of decision makers may be viewed as analogous to an observation in PCA, and each of the m decision makers may be viewed as a variable or as one dimension of the data set. The decision makers themselves are individually different, and thus one would expect to see some variability in the numerical scores assigned by them to a given comparison. The objective of the principal components analysis would then be to determine the (uncorrelated) principal components that capture the majority of the variability among the judges.

Our general approach is as follows: we first replace all of the original comparison values with their logarithms. This converts the discrete values that comprise the Fundamental Scale of Absolute Numbers (Saaty, 1980; 1990) to continuous values, facilitating principal component calculation. We then convert back to the ratio scale at the end of our method. This also reduces the problem to finding decision maker weights for use with a weighted GM. To see this, suppose decision maker k is assigned a weight w_k where $w_k \in (0,1)$ and $\sum_{k=1..m} (w_k)=1$. Further let a_{ij}^k represent the value from Saaty's Fundamental Scale of Absolute Numbers (Saaty 1980; 1990) chosen by decision maker k in comparing factor i with factor j. Let us also denote by a_{ij} the weighted geometric mean of these values across all of the judges, i.e., $a_{ij} = \prod_{k=1}^{m} (a_{ij}^k)^{w_k}$. First, note that if a_{ij} is used as the final synthesized value for the comparison between factors i and j, then as desired in the AHP, $a_{ji} = \prod_{k=1}^{m} (1/a_{ij}^k)^{w_k} = (1/a_{ij})$. Now suppose that we replace a_{ij}^k with $log(a_{ij}^k)$. If we compute the weighted arithmetic mean of these values across the m decision makers we obtain $\sum_{k=1}^{m} w_k log\,(a_{ij}^k) = log\left(\prod_{k=1}^{m} (a_{ij}^k)^{w_k}\right) = log\ (a_{ij})$, and the corresponding value in the transposed entry is equal to $log\ (1/a_{ij})$. Exponentiating these values will thus yield synthesized values that are identical to those from the weighted GM.

Therefore, we compute the principal components for the original matrix of comparisons (in their logarithms), and restrict ourselves to the first principal component, which is the eigenvector corresponding to the largest eigenvalue of the corresponding covariance matrix. By definition, this *m*-vector captures the majority of the variance, and we normalize and use this as the vector of weights to develop final aggregated values for the numerical comparisons, which in turn are then used to develop the final set of priorities in the AHP. Note that this does not involve any distributional assumptions, so the approach may be used with any dispersed data set. The approach may also be viewed as a way of obtaining an appropriate set of weights when developing the priority vector in the context of group aggregation via the weighted geometric mean regardless if the pairwise comparisons are dispersed or not. In the next section, we illustrate the PCA-based approach with a numerical example from an actual application. We then further examine this approach by studying how the weights behave with respect to the magnitude of the disagreement among the judges, and show that as the dispersions tend to zero, the weights from the PCA tend to equal values for all the judges, so that the weighted GM converges to the regular (unweighted or raw) GM.

4.0 ILLUSTRATION OF PROCEDURE

We illustrate how the approach works using an example from the nuclear power generation sector. The data relates to the development of a methodology for managing spare parts at a Fortune 200 company operating in this sector, which was the motivation behind the development of this approach (Scala, Rajgopal, and Needy, 2014). The AHP has been used to classify inventory and spare parts in other situations as well; examples of such studies include Molenaers et al. (2012) and Lolli, Ishizaka, and Gamberini (2014). One phase of our study involved collecting data from subject matter experts (SMEs) regarding the current spare parts ordering process. The process was mapped in an influence diagram, which included 34 influences. These influences were arranged into seven focus areas: timeliness of work order, part failure, vendor availability, part usage in plant, preventive maintenance schedule, outage usage, and cost consequences. SMEs were asked to perform pairwise comparisons of the aspects of the existing system by focus area; each SME only provided comparisons for areas in which he/she is an expert. In total, and to capture a holistic data set, twenty-five SMEs were asked to perform the pairwise comparisons, with five SMEs assigned to each set of influences. SMEs from more than one company location (generation plants and corporate offices) were used because each facility handles spares within its plant or office. In the larger study, the authors had the aim of developing a spares methodology to implement at all of the company's plants. Hence, a variety of inputs from multiple SMEs from various backgrounds and focus areas were included. Employee knowledge was essential to the larger model. One employee's perspective was simply not enough due to company sub-cultures and the systematic nature of the spare parts process. The SMEs were geographically dispersed across three of the company's nuclear generation plants and its corporate offices; these facilities are located over two states. Various impediments prevented the SMEs from gathering in one central location, so judgments were elicited individually, which also helped to eliminate bias from group-think or an overly assertive SME. Furthermore, not all SMEs participated in the pairwise comparisons for every set, as expertise varied among the decision makers. A full discussion of the elicitation process and influence sets, including pairwise comparison data for all SMEs, can be found in Scala (2011) and Scala, Needy, and Rajgopal (2010).

For any pairwise comparison, the numerical value assigned was allowed to take on one of nine values (1/9, 1/7, 1/5, 1/3, 1, 3, 5, 7, 9) from Saaty's Fundamental Scale of Absolute Numbers traditionally used in the AHP (Saaty, 1980; 1990). The scale is monotonic in the sense of strength of preference for one alternative over the other, so that a value of 1/9 would represent one end of the opinion spectrum and a value of 9 the other end. Thus if the decision makers are consistent they would assign values for the comparison that are close to each other, while a divergence of opinion would result in values that cover a wider range.

To illustrate our approach, consider the following AHP judgments that were elicited from all five SMEs for one set of influences that related to the preventive maintenance schedule at the company (Scala 2011). There were f=4 alternatives being compared against each other, and the 5-vector in row i and column j (i≠j) in the matrix below represents the numerical values assigned by the five judges to the comparison between alternatives i and j, (for each of a total of 6 pairwise comparisons):

$$\begin{bmatrix} 1 & (3,\frac{1}{3},5,5,5) & (7,\frac{1}{5},\frac{1}{3},5,\frac{1}{5}) & (1,\frac{1}{5},\frac{1}{7},5,\frac{1}{7}) \\ & 1 & (3,\frac{1}{3},\frac{1}{5},3,\frac{1}{3}) & (1,\frac{1}{5},\frac{1}{5},1,\frac{1}{5}) \\ & & 1 & (\frac{1}{5},\frac{1}{3},\frac{1}{7},5,\frac{1}{5}) \\ & & & 1 \end{bmatrix}$$

The dispersion test of Saaty and Vargas was performed on each group of five pairwise comparison judgments above. All six groups failed the test due to excessive dispersion. Each of the decision makers was contacted via follow-up individual interviews, when a few judgments were revised but most were not. Most SMEs did not want to revise their judgments because they felt that what was initially provided was an accurate assessment of the spare parts process at their respective company location. To avoid bias and data integrity issues, the decision makers were not forced to revise their judgments; we were seeking a more accurate representation of the current process as data inputs to the full study. The revised pairwise comparison matrix for the preventive maintenance influences is shown below:

$$\begin{bmatrix} 1 & (3,\frac{1}{3},5,5,1) & (7,\frac{1}{5},\frac{1}{3},5,\frac{1}{5}) & (1,\frac{1}{5},\frac{1}{7},5,\frac{1}{7}) \\ & 1 & (3,\frac{1}{3},\frac{1}{5},3,\frac{1}{3}) & (1,\frac{1}{5},\frac{1}{3},1,\frac{1}{5}) \\ & & 1 & (\frac{1}{5},\frac{1}{3},\frac{1}{5},5,\frac{1}{5}) \\ & & & 1 \end{bmatrix}$$

The dispersion test was rerun, and once again all sets of comparisons failed the test. At this stage we treated the comparison matrix as final and our PCA-based approach to aggregation was used. We start the aggregation with the following "comparison matrix" $\boldsymbol{A}$ obtained from the data above; each row denotes one specific pairwise comparison and each column denotes one specific judge:

$$\boldsymbol{A} = \begin{bmatrix} 3 & 1/3 & 5 & 5 & 1 \\ 7 & 1/5 & 1/3 & 5 & 1/5 \\ 1 & 1/5 & 1/7 & 5 & 1/7 \\ 3 & 1/3 & 1/5 & 3 & 1/3 \\ 1 & 1/5 & 1/3 & 1 & 1/5 \\ 1/5 & 1/3 & 1/5 & 5 & 1/5 \end{bmatrix}$$

We define the transformed matrix $\boldsymbol{X}$, each entry of which is the logarithm of the corresponding entry in $\boldsymbol{A}$. Without going through the details of the computations, it may be verified using any standard computational software package that the first principal component (the unit eigenvector corresponding to the largest eigenvalue of the covariance matrix for $\boldsymbol{X}$) is the vector $[0.5427\ \ 0.0446\ \ 0.7424\ \ 0.0701\ \ 0.3840]^{T}$. Note that this vector has its *l*-2 norm equal to 1; because the AHP approach uses the *l*-1 norm, we simply square each element (so that they sum to 1.0) to obtain the final vector of "weights" to be given to each judge as $\boldsymbol{w} = [0.2946\ \ 0.0020\ \ 0.5511\ \ 0.0049\ \ 0.1475]^{T}$. It may also be noted that the vector of sample variances of the logarithms of the judgments by the five decision makers is given by $[1.5442\ \ 0.0783\ \ 1.6912\ \ 0.4204\ \ 0.4868]^{T}$; normalizing this yields the vector $[0.3658\ \ 0.0185\ \ 0.4007\ \ 0.0996\ \ 0.1153]^{T}$. Comparing this with the vector $\boldsymbol{w}$ shows that while the weights are not the same, the relative magnitudes bear some resemblance (e.g., judges 2 and 4 get low weights, while judge 3 gets the highest weight). Furthermore, company leadership reacted positively to the weights for each judge. Not all judges knew each other well, so it is difficult to assess individual satisfaction for each judge. However, leadership had a high level understanding of each location and its corresponding workforce.

Given the vector of weights $\boldsymbol{w}$, the aggregated value for each comparison would be the weighted geometric mean of the original judgments and is given by $\boldsymbol{z} = [3.3746\ \ 0.7673\ \ 0.2581\ \ 0.4857\ \ 0.4292\ \ 0.2034]$. (Alternatively we could find $\boldsymbol{z'}=\boldsymbol{Xw}$, and then use $\boldsymbol{z}=\exp(\boldsymbol{z'})$ to get these same final values.) Thus the set of synthesized judgments for computing the AHP priorities is as shown below in Table 1:

Insert Table 1 Here

Table 1: Matrix of synthesized judgments for computing AHP priorities

The row and column labels in Table 1 reference the alternatives or criteria that are being pairwise compared. For example, position *(1,2)* in the table is the synthesized judgment for the pairwise comparison between the first and second attribute or criterion. Using the approach of Saaty (1980; 1990), the (normalized) principal right eigenvector for this matrix yields the final priority vector for the AHP corresponding to the four factors as $\boldsymbol{v}^{\boldsymbol{P}} = [0.3490\ \ 0.0061\ \ 0.6292\ \ 0.0157]^T$ for the AHP.

Note that if instead, we had used the raw geometric mean of the five judgments for each pairwise comparison, we obtain the vector $\boldsymbol{z} = [1.9037\ 0.8586\ 0.4592\ 0.7248\ 0.4217\ 0.4217]^T$ so that the synthesized symmetric matrix for computing the priorities would yield Table 2:

Insert Table 2 here

Table 2: Matrix of synthesized judgments for computing AHP priorities using the GM

This yields the priority vector $\boldsymbol{v}^{\boldsymbol{G}} = [0.3743\ \ 0.1475\ \ 0.4353\ \ 0.0429]^T$.

The less dispersion there is among the judges, the closer these two different vectors $\boldsymbol{v}^{\boldsymbol{G}}$ and $\boldsymbol{v}^{\boldsymbol{P}}$ are to each other. For instance, suppose the original comparison matrix was given by

$$\boldsymbol{A'} = \begin{bmatrix} 3 & 3 & 5 & 5 & 5 \\ 1/7 & 1/5 & 1/3 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/7 & 1/5 & 1/7 \\ 1/3 & 1/3 & 1/5 & 1/3 & 1/3 \\ 1/5 & 1/5 & 1/3 & 1/3 & 1/5 \\ 1/5 & 1/3 & 1/5 & 1/5 & 1/5 \end{bmatrix}$$

Several of the numbers were altered to capture a situation where there is still some variability in $\boldsymbol{A'}$ but significantly less than in the original matrix $\boldsymbol{A}$. In this case, the weights using the PCA approach are given by $[0.1705\ \ 0.1480\ \ 0.2209\ \ 0.2168\ \ 0.2438]^T$

leading to the synthesized judgments [4.2493 0.2114 0.1711 0.2978 0.2501 0.2157]. On the other hand the raw GM yields the synthesized judgment [4.0760 0.2071 0.1748 0.3010 0.2453 0.2215]. The two corresponding final priority vectors are given by $\boldsymbol{v}^{\boldsymbol{P}}$ = [0.6165 0.0649 0.3096 0.0090]^T for the PCA based approach and $\boldsymbol{v}^{\boldsymbol{G}}$= [0.6199 0.0773 0.2939 0.0090]^T for the approach based on the raw GM. As one can observe, these are very similar to each other. In particular, consider the limiting case when there is *no* dispersion and the decision makers are perfectly consistent in their judgments, so that each n-dimensional column vector in $\boldsymbol{A}$ is identical (say $\boldsymbol{y}$) with all judges assigning exactly the same scores to any comparison. In this situation the first principal component is a vector $\boldsymbol{w}$ of dimension m, with each entry equal to $1/m$, i.e., each of the m judges is assigned the same weight of $1/m$. Thus in this case the weighted GM is identical to the raw GM, and the final priority vectors $\boldsymbol{v}^{\boldsymbol{P}}$ and $\boldsymbol{v}^{\boldsymbol{G}}$ are identical.

5.0 LIMITING BEHAVIOR OF PCA-BASED APPROACH

In this section we describe an empirical study, with the objective of understanding the limiting behavior of the PCA based approach for aggregation as diversity of opinion among the decision makers decreases and to demonstrate that it converges uniformly to aggregation by the GM. The data for the study is generated from a simulation model. Before describing the range of parameter values studied, we first describe how "diversity of opinion" (i.e., dispersion) is captured by the study. Suppose we number each of the nine values on Saaty's Fundamental Scale as a "unit" as shown in Table 3 below:

Insert Table 3 here

Table 3: Correspondence of units to Saaty's Fundamental Scale of Absolute Numbers

To capture variation among the decision makers, we then define the "scale spread" (δ) as the range of units covered by the values assigned by different decision makers in the group. So, for example, if δ=2 all decision makers chose one of two adjacent units on the scale. On the other hand, if δ=5 they all chose values from among a set of 5 consecutive units on the scale (e.g., {1/7, 1/5, 1/3, 1, 3} or {1/3, 1, 3, 5, 7} or {1, 3, 5, 7, 9}). In the two extreme cases, if δ=1, then all decision makers are perfectly consistent and chose the same value, while if δ=9, their values ranged across the entire spectrum of possible values (from 1/9 through 9). In order to conduct our study we vary three different parameters as follows:

a. The number of decision makers: m=4,5,6,7,8,10,15,25,50
b. The number of factors/alternatives being compared with each other: f=3,4,5,6,7,8, and
c. The scale spread δ=2,3,4,5,6,7,8,9

Note that if f alternatives are being compared with each other, the actual number of pairwise comparisons being made is $n=f(f-1)/2$; this would also be the number of rows in a "comparison matrix" $\boldsymbol{A}$. The comparison matrix is constructed with each row corresponding to a specific pairwise comparison and each column corresponding to a decision maker in the group. The objective is to study the difference between the priority vectors from a PCA-based weighted geometric mean approach and the raw GM-based approach on the same set of judgments when there are different levels of disagreement among the decision makers (as measured by δ), across different levels of m and f. Of course, if δ=1 the priority vectors are always identical.

For a given combination of values for m, f and δ, a set of m judgment values from Table 3 was randomly generated for each of the $n=f(f-1)/2$ pairwise comparisons, while ensuring that these values always range over a set of δ adjacent units on the scale. It was also ensured that both extremes in the range were definitely selected. The specific set of units spanned for each individual comparison were also randomly selected and could be different (e.g., with $\delta=3$, for comparing A vs. B the units may correspond to scale values {3, 5, 7} while for comparing C vs. D the values might be {1/3, 1, 3}). The corresponding priority vectors using the PCA approach ($\boldsymbol{v}^{\boldsymbol{P}}$) and the GM approach ($\boldsymbol{v}^{\boldsymbol{G}}$) were then computed. The vector $\boldsymbol{v}^{\boldsymbol{P}}$ was computed using the approach as described in Section 3.2, and $\boldsymbol{v}^{\boldsymbol{G}}$ was computed via the traditional geometric mean AHP judgment aggregation approach. A full factorial experiment yields a total of $|m|\cdot|f|\cdot|\delta|$= 9*6*8= 432 unique combinations, and for each combination we ran 10,000 simulations and computed the mean and standard deviation of the distances between the two vectors across these replicates. Note that in each replicate there are $n=f(f-1)/2$ separate pairwise comparisons.

Before comparing vectors $\boldsymbol{v}^{\boldsymbol{P}}$ and $\boldsymbol{v}^{\boldsymbol{G}}$, we first examine the circumstances under which a raw GM based approach for aggregating group judgments would be inappropriate. To see this, for each combination of inputs (m, f and δ), we also computed the percentage of the (10,000*n) comparisons in the simulation where the GM-based approach would fail the dispersion test of Saaty and Vargas (2007) at a p value of 0.05; suppose we use $\alpha_{mf\delta}$ to denote this percentage. First, note that once the m and δ are fixed, f only determines how many comparisons are being made; the individual comparisons are similar in that they have the same spread across the same number of judges. Thus we average the percentages across all values of f for a given combination of m and δ, i.e., we compute $\alpha_{m\cdot\delta} = \{\sum_{f=3,\ldots,8}\alpha_{mf\delta}\} \div 6$. Based on this, Table 4 provides simulated estimates for each m, the probability $(\alpha_{m\cdot\delta})/100$ that the dispersion test fails when there is a scale spread of δ. Combinations that yield a value of 0.05 or lower are shaded.

Insert Table 4 here

Table 4: The probability the dispersion test fails with a scale spread of δ

Figure 1 plots the results to better visualize them, and each line represents a given value of m. As the results indicate, when there are more decision makers the test tolerates more divergence of opinion amongst them. When m is very large (25 or 50) a scale spread of even 4 or 5 units seems tolerable, while with very few of judges (m= 4 or 5) even a small spread of 3 units might be unacceptable. In general, these results suggest the levels of disagreement that would invalidate the use of the unweighted GM with different numbers of decision makers. This extends the work of Saaty and Vargas (2007), as they do not provide formal guidance (nor does the literature) on how many sets of pairwise comparisons must fail before the unweighted geometric mean is no longer appropriate.

Insert Figure 1 here

Fig. 1 Plot of the probability the dispersion test fails with a scale spread of δ

In comparing the vectors $\boldsymbol{v}^{\boldsymbol{P}}$ and $\boldsymbol{v}^{\boldsymbol{G}}$ it is natural to use some sort of a distance measure. However, it should be kept in mind that the priority vectors do not have the same dimension; they belong to $\boldsymbol{R}_{+}^{\boldsymbol{f}}$ (and have an ℓ-1 norm equal to 1). Therefore a distance measure (d) between the two vectors should account for the size of f. We use the following adjusted version of the ℓ-∞ norm: $d = \frac{\{Max_{i=1,\dots,f}|v_i^P - v_i^G|\}}{(1/f)} * 100.$ The denominator represents the average priority value for a set of f factors, so that d represents the distance constituted by the l -∞ norm as a percentage of this average.

Figure 2 displays a series of plots, one for each value of f. In each graph we plot the distance d as a function of δ for each of the nine values of m that we considered. To maintain visual consistency, all plots have the same scale. It can be observed that as δ approaches 1 (i.e., we approach complete consistency among the judges) the distance d converges uniformly to zero in all cases. Furthermore, for any fixed value of f and δ, the distance between $\boldsymbol{v^G}$ and $\boldsymbol{v^P}$ is always smaller when the number of judges is larger. This is intuitive; for the same level of dispersion, the effect of one individual judge is less pronounced with more judges. Second, for any fixed f, differences between $\boldsymbol{v^G}$ and $\boldsymbol{v^P}$ generally do not depend strongly on the number of judges when there is relatively less dispersion (say, δ=4 or lower). However, the difference between the priority vectors as a function of m becomes more pronounced when the dispersion is relatively large (say, δ=6 or higher). Finally, the distance between the priority vectors rises monotonically with dispersion in most cases.

Insert Figure 2 here

Fig. 2 Distance d as a function of δ for each value of f

In summary, the results of the simulation show that the final priority vector found by aggregating the dispersed judgments using the PCA-based method, with weights for every decision maker found via the first principal component vector, generalizes to the priority vector found when the judgments are aggregated using a traditional geometric mean with the decision makers not dispersed. This shows the PCA-based approach is no different from the approach using the raw or unweighted GM with the added capability of identifying an approach to the aggregation of judgments if the dispersion test does not pass. Furthermore, because the PCA-approach generalizes to the raw GM, it can be used to find weights for decision makers any time a weighted geometric mean is desired.

6.0 SUMMARY

This paper addresses the issue of synthesizing the AHP judgments of multiple decision makers when the judgments do not pass the dispersion test and the decision makers are unwilling or unable to revise their judgments. In these cases, significant diversity of opinion exists amongst the decision makers. Prior work on this topic has focused mainly on identifying such situations but provides no guidance on how to deal with this issue, other than asking the decision makers to reconsider their judgments. Simply discarding a judgment that is not consistent with the majority is also not a satisfactory solution because a diverse set of perspectives may need to be obtained, as was the case in the study that motivated this research. While judgments of all the decision makers should be considered, there is no reason in general why all decision makers need be given the same degree of importance when synthesizing their judgments. This leads to the natural choice of a weighted geometric mean, which preserves the axioms that are critical to the AHP. However, the research literature does not provide any guidance on how the weights can be selected in an objective fashion, especially in the case of excessive dispersion. Motivated by a supply chain problem, we propose an aggregation approach based on PCA that treats the judges as variables in a set of comparisons and uses the first principal component (which captures the maximum amount of variability) to arrive at weights for each decision maker. The use of PCA flushes out importance and relevance of decision makers by accounting for dispersion. A detailed simulation provided guidance on when this approach might be preferred over a simple geometric mean and showed that the final priority vector from this approach converges uniformly to the one obtained from a simple geometric mean as the diversity of opinion among the judges tends to zero.

REFERENCES

Aczél J, Saaty TL (1983) Procedures for synthesizing ratio judgments. J Math Psychol 27:93-102.

Aczél J, Alsina C (1987) Synthesizing judgments: A functional equations approach. Math Modelling 9:311-320.

Anderson TW (2003) An introduction to multivariate statistical analysis 3rd ed. Wiley-Interscience, Hoboken

Arias-Nicolás JP, Pérez CJ, Martín J (2008) A logistic regression-based pairwise comparison method to aggregate preferences. Group Decis Negotiat 17:237-247.

Armacost RL, Hosseini JC, Pet-Edwards J (1999) Using the Analytic Hierarchy Process as a two-phase integrated decision approach for large nominal groups. Group Decis Negotiat 8:535-555.

Basak I (1988) When to combine group judgments and when not to in the Analytic Hierarchy Process: A new method. Math Comput Modelling 10:395-404.

Basak I, Saaty T (1993) Group decision making using the Analytic Hierarchy Process. Math Comput Modelling 17:101-109.

Cho YG, Cho KT (2008). A loss function approach to group preference aggregation in the AHP. Comput Operations Res 35: 884-892.

Dunteman GH (1989) Principal components analysis. Sage Publications, Newbury Park

Dyer RF, Forman EH (1992) Group decision support with the Analytic Hierarchy Process. Decis Support Syst 8:99-124.

Escobar MT, Moreno-Jiménez JM (2007) Aggregation of individual preference structures in AHP-group decision making. Group Decis Negotiat 16:287-301.

Forman E, Peniwati K (1998) Aggregating individual judgments and priorities with the Analytic Hierarchy Process. Eur J Operational Res 108:165-169.

Huang Y, Liao J, Lin Z (2009) A study on aggregation of group decisions. Syst Res Behav Sci 26:445-454.

Ishizaka A, Labib A (2011a) Review of the main developments in the Analytic Hierarchy Process. Expert Syst Appl 38:14336-14345.

Ishizaka A, Labib A (2011b) Selection of new production facilities with the Group Analytic Hierarchy Process Ordering method. Expert Syst Appl 38:7317-7325.

Johnson RA, Wichern DW (2007) Applied multivariate statistical analysis 6th ed. Pearson Prentice Hall, Upper Saddle River

Jolliffe IT (2002) Principal component analysis 2nd ed. Springer-Verlag, New York

Korpela J, Lehmusvaara A, Nisonen J (2007) Warehouse operator selection by combining AHP and DEA methodologies. Int J Prod Econ 108:135-142.

Lai VS, Wong BK, Cheung W (2002) Group decision making in a multiple criteria environment: A case using the AHP in software selection. Eur J Operational Res 137: 134-144.

Liberatore MJ, Nydick RL (1997) Group decision making in higher education using the Analytic Hierarchy Process. Res High Educ 38:593-614.

Liu FF, Hai HL (2005) The voting Analytic Hierarchy Process method for selecting supplier. Int J Prod Econ 97:308-317.

Lolli F, Ishizaka A, Gamberini R (2014) New AHP-based approaches for multi-criteria inventory classification. Int J Prod Econ 156:62-74.

Molenaers A, Baets H, Pintelon L, Waeyenbergh G (2012) Criticality classification of spare parts: A case study. Int J Prod Econ 140:570-578.

Pedrycz W, Song M (2011) Analytic Hierarchy Process (AHP) in group decision making and its optimization with an allocation of information granularity. IEEE Transactions Fuzzy Syst 19:527-539.

Ramanathan R (2007) Supplier selection problem: Integrating DEA with the approaches of total cost of ownership and AHP. Supply Chain Manag Intl J 12:258-261.

Ramanathan R, Ganesh LS (1994) Group preference aggregation methods employed in AHP: An evaluation and an intrinsic process for deriving members' weightages. Eur J Operational Res 79:249-265.

Saaty TL (1980) The Analytic Hierarchy Process: Planning, priority setting, resource allocation. McGraw-Hill, New York

Saaty TL (1990) How to make a decision: The Analytic Hierarchy Process. Eur J Operational Res 48:9-26.

Saaty TL (2013) The modern science of multicriteria decision making and its practical applications: The AHP/ANP approach. Operations Res 61:1101-1118.

Saaty TL, Vargas LG (2005) Dispersion of group judgments. Proc ISAHP 2005.

Saaty TL, Vargas LG (2007) Dispersion of group judgments. Math Comput Modelling 46:918-925.

Saaty TL, Vargas LG (2012) The possibility of group choice: Pairwise comparisons and merging functions. Soc Choice Welf 38:481-496.

Scala NM (2011) Spare parts management for nuclear power generation facilities. Dissertation, University of Pittsburgh

Scala NM, Needy KL, Rajgopal J (2010) Using the Analytic Hierarchy Process in group decision making for nuclear spare parts. Proc 31st Annu ASEM Natl Conf.

Scala NM, Rajgopal J, Needy KL (2014) Managing spare parts inventories: A data driven methodology. IEEE Transactions Eng Manag 61:28-37.

Subramanian N, Ramanathan R (2012) A review of applications of Analytic Hierarchy Process in operations management. Int J Prod Econ 138:215-241.

Van den Honert RC (1998) Stochastic group preference modelling in the multiplicative AHP: A model of group consensus. Eur J Operational Res 110:99-111.

Van den Honert RC (2001) Decisional power in group decision making: A note on the allocation of group members' weights in the multiplicative AHP and SMART. Group Decis Negotiat 10:275-286.

Van den Honert RC, Lootsma FA (1996) Group preference aggregation in the multiplicative AHP: The model of the group decision process and Pareto optimality. Eur J Operational Res 96:363-370.

Vargas L (1997) Comments on Barzilai and Lootsma, why the multiplicative AHP is invalid; A practical counterexample. J Multi-criteria Decis Anal 6:169-170.

Wei C, Chien C, Wang MJ (2005) An AHP-based approach to ERP system selection. Int J Prod Econ 96:47-62.

Xu Z, Yager RR (2010) Power-geometric operators and their use in group decision making. IEEE Transactions Fuzzy Syst 18:94-105.

Group Decision Making with Dispersion in the Analytic Hierarchy Process

Table 1

	1	2	3	4
1	1	3.3746	0.7673	0.2581
2	1/3.3746	1	0.4857	0.4292
3	1/0.7673	1/0.4857	1	0.2034
4	1/0.2581	1/0.4292	1/0.2034	1

Table 1: Matrix of synthesized judgments for computing AHP priorities

Table 2

	1	2	3	4
1	1	1.9037	0.8586	0.4592
2	1/1.9037	1	0.7248	0.4217
3	1/0.8586	1/0.7248	1	0.4217
4	1/0.4592	1/0.4217	1/0.4217	1

Table 2: Matrix of synthesized judgments for computing AHP priorities using the GM

Table 3

Unit	Value
1	1/9
2	1/7
3	1/5
4	1/3
5	1
6	3
7	5
8	7
9	9

Table 3: Correspondence of units to Saaty's Fundamental Scale of Absolute Numbers

Table 4

δ	No. of judges (m) 4	5	6	7	8	10	15	25	50
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.133	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.421	0.327	0.217	0.125	0.102	0.060	0.008	0.001	0.000
4	0.758	0.584	0.497	0.407	0.359	0.252	0.139	0.033	0.008
5	1.000	0.937	0.818	0.747	0.672	0.546	0.405	0.241	0.072
6	1.000	1.000	0.977	0.936	0.927	0.874	0.761	0.589	0.395
7	1.000	1.000	0.997	0.980	0.974	0.964	0.940	0.905	0.838
8	1.000	1.000	1.000	0.993	0.987	0.986	0.985	0.986	0.991
9	1.000	1.000	1.000	0.997	0.993	0.992	0.995	0.998	1.000

Table 4: The probability the dispersion test fails with a scale spread of δ

Group Decision Making with Dispersion in the Analytic Hierarchy Process

Figure 1

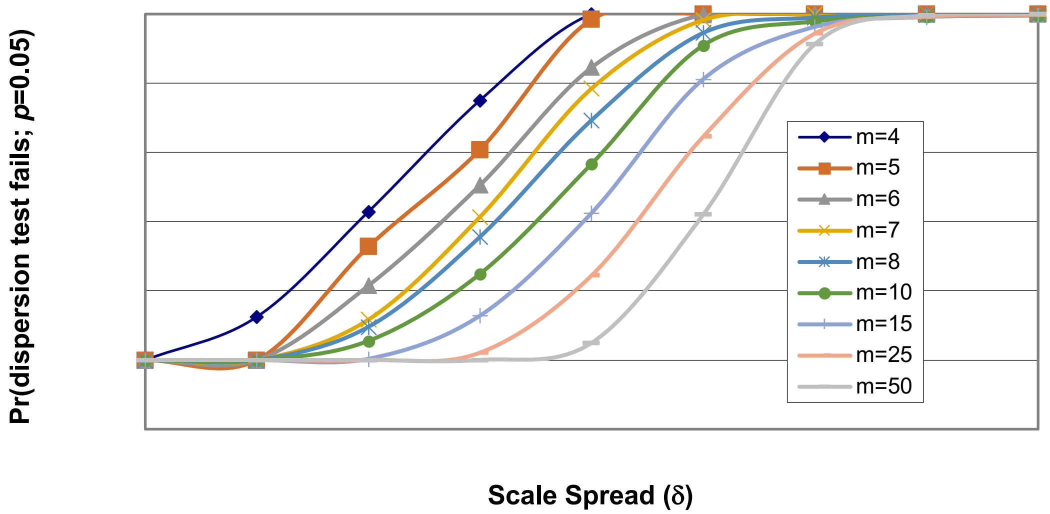

Fig. 1 Plot of the probability the dispersion test fails with a scale spread of δ

Group Decision Making with Dispersion in the Analytic Hierarchy Process: Figure 2

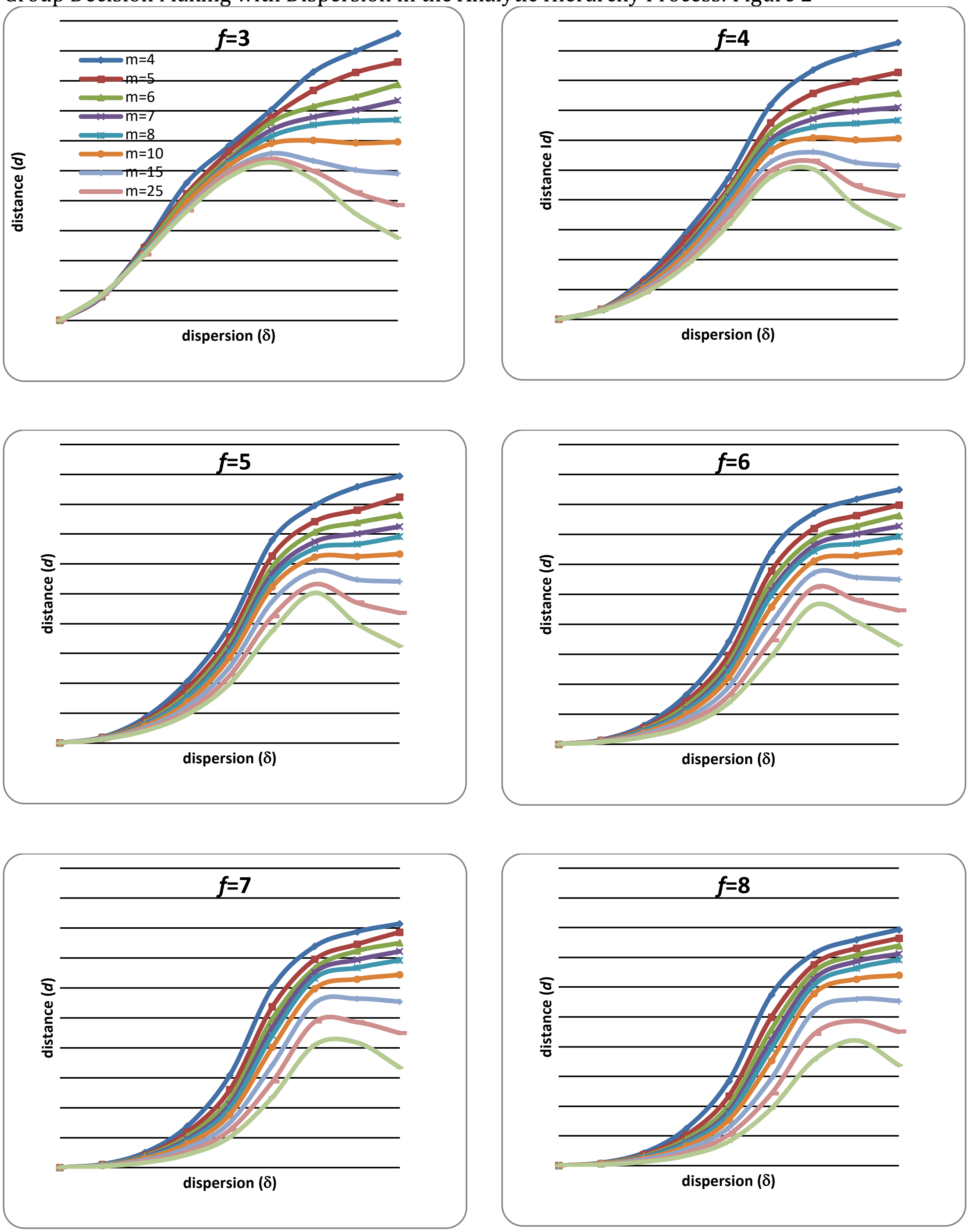

Fig. 2 Distance d as a function of δ for each value of f